Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of	)	
	)	
Establishing the Digital Opportunity Data Collection	)	WC Docket No. 19-195
	)	
	)	
Modernizing the FCC Form 477 Data Program	)	WC Docket No. 11-10
	)	

REPLY COMMENTS
of
MICROSOFT CORPORATION

MICROSOFT CORPORATION

Paula Boyd, Senior Director
U.S. Government and Regulatory Affairs
901 K Street, N.W., 11th Floor
Washington, DC 20001
202-263-5946

David A. LaFuria
Lukas, LaFuria, Gutierrez & Sachs, LLP
8300 Greensboro Drive, Suite 1200
Tysons, VA 22102
703-584-8678

October 7, 2019

TABLE OF CONTENTS

SUMMARY

Microsoft commends the Commission for adopting a new Digital Opportunity Data Collection. Accurate measurements of broadband availability and usage will assist Congress and the Commission in making sound legislative and policy decisions that can drive efficient investments needed to close the Digital Divide and ensure that Americans living in rural areas have access to broadband that is reasonably comparable to those living in our urban areas.

Microsoft suggests a number of steps the Commission can take to improve the quality of its broadband availability data.

- Align the fixed broadband availability instruction on Form 477 to match the newly adopted DODC instruction.
- Begin an orderly transition toward sunsetting Form 477 availability data collection once the DODC is successfully implemented, while retaining Form 477 for the collection of other data.
- Improve the accuracy of the DODC data by:
    - Employing machine learning tools to efficiently and accurately analyze the DODC data that the Commission will receive.
    - Enabling the collection and analysis of at least three types of crowdsourced data – availability data, speed test and/or usage data, and subscription data. Recent Microsoft research has revealed that usage data is an excellent predictor of broadband availability, enabling the FCC and providers to identify outliers in their availability data reporting.
    - Enabling full transparency of all FCC-collected data, enabling the public and other stakeholders to identify discrepancies and rapidly improve data quality.
    - Implement a broadband fabric and a lookup tool so that all stakeholders have a common view of broadband serviceable locations.

With respect to the submission of broadband data, polygons should be no smaller than the building level, and in the aggregate, every building in the United States should be represented. All polygons must be anchored onto the same fabric and use the same geocoding reference points so that a single standard base map is used to count serviceable locations and all maps and data sets must contain uniform characteristics.

Finally, Microsoft supports reasonable accountability measures that deter willful, intentional, or reckless submission of data that significantly overstates broadband coverage.

Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of	)	
	)	
Establishing the Digital Opportunity Data Collection	)	WC Docket No. 19-195
	)	
	)	
Modernizing the FCC Form 477 Data Program	)	WC Docket No. 11-10
	)	

REPLY COMMENTS
of
MICROSOFT CORPORATION

Microsoft Corporation ("Microsoft") provides the reply comments below in response to the Second Further Notice of Proposed Rulemaking, adopted by the Commission in the above-captioned dockets.[1]

I. Introduction

Microsoft supports the Commission's continuing efforts to improve its data collection process, to increase visibility into broadband deployment throughout the nation, and to increase the efficiency of its universal service support investments. The newly adopted Digital

[1] *Establishing the Digital Opportunity Data Collection*, *et al*., WC Docket No. 19-195, *et al*., Report and Order and Second Further Notice of Proposed Rulemaking, FCC 19-79 (rel. Aug. 6, 2019) ("*Report and Order*" or "*Second Notice*"). Reply comments in this proceeding are due not later than October 7, 2019. FCC, *Establishing the Digital Opportunity Data Collection and Modernizing the FCC Form 477 Data Program*, 84 Fed. Reg. 43764 (Aug. 22, 2019).

Opportunity Data Collection ("DODC") represents a significant step forward. Having an accurate picture of where broadband is available, and where it is not, is critical to accurately crafting broadband policies and targeting investments of scarce universal service support funds, as well as other federal and state broadband investments; providing accurate reports to Congress required by Section 706 of the Telecommunications Act of 1996; and improving the public's understanding of broadband availability.

At the outset, Microsoft suggests a number of steps the Commission can take to improve the quality of its broadband availability data.

- Align the fixed broadband availability instruction on Form 477 to match the newly adopted DODC instruction.
- Begin an orderly transition to sunset Form 477 availability data collection once the DODC is successfully implemented, while retaining Form 477 for the collection of other data.
- Improve the accuracy of the Commission's DODC data by:
    - Employing machine learning tools to efficiently and accurately analyze the DODC data that the Commission will receive.
    - Enabling the collection and analysis of at least three types of crowdsourced data – availability data, speed test and/or usage data, and subscription data. Recent Microsoft research has revealed that usage data is an excellent predictor of broadband availability, enabling the FCC and providers to identify outliers in their availability data reporting.
    - Enabling full transparency of all FCC-collected data, enabling the public and other stakeholders to identify discrepancies and rapidly improve data quality.
    - As detailed below, implementing a broadband fabric and a lookup tool so that all stakeholders have a common view of broadband serviceable locations.

It is important for the Commission to address these issues aggressively now, because the practical effects of not doing so are significant. Without broadband, consumers and

businesses cannot access economic, educational, healthcare, social and other opportunities, many of which are enabled by cloud products and services provided by Microsoft and others. Increasing all of these opportunities aligns with Microsoft's mission to "empower every person and organization on the planet to achieve more."[2] To accomplish this mission, a critical first step is accurately measuring broadband availability.

This proceeding, coupled with the Commission's renewed focus on effective data practices, as evidenced by the Commission's creation of its Office of Economics and Analytics ("OEA"), as well as the general emergence of data analytics and predictive analytical tools, offer a unique opportunity to strengthen the Commission's data analytics foundation.

II. The Commission Should Use a Variety of Data Sets and Leverage Machine Learning Tools to Improve the Accuracy of Broadband Availability Mapping.

The Commission is correct that the new DODC "represents a unique opportunity for integrating related but distinct data resources to produce a unified picture of broadband data."[3] Microsoft offers unique expertise in enabling and managing large scale data and measurement systems. We do this for our business and with our customers worldwide. We also have world class data scientists who deploy Machine Learning (ML) to help customers and Microsoft scale. In this space, we have identified four key components that any ML system should have to be successful. Thus, we recommend the following:

---

[2] Indeed, Microsoft is pursuing this goal worldwide. *See* https://www.microsoft.com/en-us/about.

[3] *Second Notice* at ¶ 84.

First, machine learning tools can analyze large data sets with accuracy and speed far greater than traditional methods. As a part of their data science work and in accordance with privacy principles, many technology companies, non-profits, and academic and research institutions leverage a large volume of anonymized data. Machine learning tools are essential to predicting problems in ways that would be difficult or impossible for a manual-only reporting system to replicate. In its DODC, the Commission will be dealing with data sets that are too large for human beings or simple data analysis tools to analyze efficiently. Machine learning models can assist in identifying anomalies in broadband availability data with far greater speed and accuracy.

Second, to leverage the power of machine learning to achieve a more accurate picture of broadband availability, a variety of data sets must be added to the model, including most importantly, usage data, which should be in the form of both subscription data (collected on FCC Form 477) and crowdsourced usage data. An initial ML model built by Microsoft to assist stakeholders in this area shows great promise at predicting inaccuracies with broadband availability data reporting. We combined Microsoft broadband usage data, US Census data, and other data, to predict outliers in the current broadband availability reporting. The model revealed that the highest predictor of availability is usage data, and thus usage is a key ingredient to any future automation via ML to understand discrepancies in availability reporting.[4] This model was demonstrated to the FCC and we are making it publicly available on

[4] Indeed, Microsoft's usage data correlates with the FCC's subscriber data at a 90% level, all the way down to the state level.

GitHub for all stakeholders to improve upon by adding other demographic data sets, such as, for example, aggregated unemployment, income levels, and crowdsourced speed tests. Adding data sets increases the likelihood that the model will produce a better understanding of issues with respect to availability reporting.

Third, crowdsourced data is important to improving the accuracy of broadband availability data. While the Commission has so far focused on crowdsourcing broadband availability data, it should also make use of third-party crowdsourced speed test, subscription data, and usage data. With these data sources, the Commission can fully triangulate issues and identify where broadband is likely to be available, or not. It is important to note that as the quantity of crowdsourced data increases, even if some of it is lower quality, it is typical in the data science world to use various techniques to maintain the quality of the data, including creating tools that filter out anomalies.

Fourth, it is important to be transparent about data that is collected. The Commission should share with the public all broadband data it collects, allowing data scientists, researchers, and the general public to identify discrepancies and contribute to the improvement of overall data quality.

In recent meetings with the Commission, Microsoft has demonstrated how a machine learning model could identify potential outliers in current broadband availability reporting.[5]

---

[5] *See* Letter from Paula Boyd to Marlene H. Dortch, WC Docket No. 19-195, WC Docket No. 11-10 (Sept. 20, 2019) ("Microsoft September 2019 ex parte"), a copy of which is attached as Exhibit A hereto; *see also,* Letter from Paula Boyd to Marlene H. Dortch, WC Docket No. 11-10 (Dec. 5, 2018) at

We were pleased by the reception we received on the prospect of the Commission using machine learning models to increase the accuracy of broadband availability reporting. We are exploring how to refine the model to make it capable of analyzing availability data at a more detailed level.[6]

Microsoft notes the Commission's decision to allow providers to check a box in their future Form 477 submissions to maintain the confidentiality of provider-specific subscription data.[7] Form 477 subscription data is highly valuable to improving insights into broadband availability. Subscription data not only shows where people have Internet access but are not adopting it, but also is an excellent data source to use to of triangulate errors in availability data. Accordingly, Microsoft recommends that aggregated subscription data that is not identifiable to a specific provider be made publicly available for inclusion in any machine learning model developed for analyzing broadband availability.

---

https://ecfsapi.fcc.gov/file/1206313012489/2018%201205%20ex%20parte%20letter%20FINAL.pdf; Letter from Paula Boyd to Marlene H. Dortch, GN Docket No. 18-238 and WC Docket No. 11-10 (Mar. 29, 2019) at https://ecfsapi.fcc.gov/file/10329025758889/2019%200329%20ex%20parte%20letter%20FINAL.pdf.

[6] Microsoft is also exploring the possibility of analyzing at the shape file level, to accommodate future DODC submissions.

[7] *Report and Order* at ¶ 37, n.94; *see also,* Appendix A, new 47 C.F.R. § 7001(d)(2)(i).

III. Given the DODC's Improvements Over Current Reporting, the Commission Should Immediately Transition the New DODC Instructions for Broadband Availability on Form 477 and, Upon Full DODC Implementation, Sunset Form 477 Broadband Availability Reporting.

The new DODC is intended to produce broadband access mapping data that is of significantly higher quality than Form 477:

> As the record in this proceeding amply demonstrates, there is a compelling and immediate need to develop granular, high-quality fixed broadband deployment data to improve our ability to target support from our Universal Service Fund (USF) programs. It has become increasingly clear that the fixed and mobile broadband deployment data collected on the Form 477 are not sufficient to support the specific imperative of our USF policy goals. We conclude that in order to continue to advance our statutory universal service obligations, it is necessary to create a new data collection, calculated to produce broadband deployment maps that will allow the Commission to precisely target scarce universal service dollars to where broadband service is lacking.[8]

The Commission seeks further comment as to whether Form 477 should sunset and on what schedule.[9] Given that the new DODC should offer tremendous improvements over current reporting mechanisms, the Commission should immediately take steps to transition the broadband availability reporting on Form 477 and ultimately sunset it in favor of DODC. Continuing to collect availability data for both Form 477 and the new DODC will be costly for providers and confusing for the public.

---

[8] *Report and Order* at ¶ 10 (footnote omitted).

[9] *Second Notice* at ¶ 135 ("We therefore seek comment on discontinuing the broadband deployment data collection that is part of Form 477 at some point after the new collection has been established").

The first step in transitioning must be aligning the Form 477 instructions for reporting broadband availability with those set forth in the new DODC.[10] This will avoid a situation where the Commission has a scorecard for its progress in increasing broadband availability that is different from the scorecard it uses to distribute universal service funds. Once the new DODC polygons come online and the broadband availability data collection is deemed by the Commission to be an improvement on Form 477, the Form 477 process should sunset for reporting broadband availability altogether.[11] The Commission should continue to collect other data on Form 477, such as for example, fixed broadband subscription data, and make that data publicly available at a minimum, to the extent such data are publicly available today.[12] Taking these steps will be especially important for small carriers with limited resources. Unless the broadband availability data collection via Form 477 is superseded, carriers will need to file two sets of broadband availability data in parallel – one for Form 477 and one for DODC – an unnecessary burden.[13]

---

[10] On this point, Microsoft has filed a petition for reconsideration in this proceeding. *See* https://ecfsapi.fcc.gov/file/109241260214870/2019%200923%20Microsoft%20Petition%20for%20Reconsideration%20-%20FINAL%20As%20Filed.pdf.

[11] A number of commenters support the sunsetting of Form 477 once the DODC process is up and running, although not all agree on precise timing, including USTelecom, ITTA, WISPA, NCTA, ACA Connects, Verizon, US Cellular, GVNW, and NRECA.

[12] *Report and Order* at ¶ 7, n.9. Retaining historical Form 477 data can be useful to identifying trends and issues in the DODC data.

[13] The Commission currently estimates the annual Form 477 burden to be 774 hours. *See* FCC Form 477 Local Telephone Competition and Broadband Reporting Instructions at 38 (Dec. 5, 2016) ("Form 477 Instructions").

The new DODC should also be used as the basis for distributing universal service funding in rural areas. Accurately targeting universal service funding has long been a core FCC objective, because areas incorrectly deemed to be served will be ineligible for funding and left behind for a long time. In addition, an improved DODC should be used to produce the Commission's periodic Section 706 reports and to develop broadband maps, which are currently based on Form 477 data that is outdated by the time that Section 706 reports that rely on the data are released.

IV. Microsoft Supports the Creation of a Broadband Fabric and a Lookup Tool.

The Commission proposes to, "create and integrate a broadband-serviceable location tool into the Digital Opportunity Data Collection."[14] It seeks comments on what kind of locations it should include as broadband-serviceable and whether to implement a lookup tool for integrating provider address data into the locations database.[15] Microsoft supports both of these initiatives.

Led by USTelecom, ITTA, and WISPA, the Broadband Mapping Coalition ("BMC") is attempting to create a broadband fabric of serviceable locations. Whether the BMC-led project is adopted, or the Commission chooses another means of creating a map of serviceable locations, it is essential for the Commission, other government agencies, and stakeholders, to be in agreement with respect to where the serviceable locations are. BMC's preliminary work

---

[14] *Second Notice*, at ¶ 101.

[15] *Id*., *see also* ¶ 108.

suggests that existing data on serviceable locations is inaccurate and the number of Americans unserved by broadband at 25/3 speed may be far greater than the 21.3 million estimated in the Commission's most recent 706 Report.[16]

If the Commission is able to combine new DODC mapping with an accurate picture of broadband serviceable locations, it will be able to provide Congress with a clearer picture of broadband deployment and availability in America that can drive future policy decisions regarding universal service funding and other federal infrastructure programs, such as the Rural Utilities Service's ReConnect Program. Improved mapping of broadband availability and serviceable locations will also help the Commission estimate what it will cost to serve those homes, businesses, and agricultural locations that remain unserved. Once estimated, the Commission can appropriately size the universal service fund, and help Congress determine whether additional funding is needed. This measurement must also be ongoing, as new technologies come online in the coming years.

Microsoft also supports the adoption of a lookup tool that permits the public to pinpoint broadband service on maps generated from the DODC. Whether the lookup tool is developed by the Commission, whether integrated from a public resource, or a commercial resource such as Bing Maps or Google Maps, the transparency of any mapping product flowing from the DODC will be improved by the addition of a lookup tool that is easy for the public to use, allowing a citizen to input an address and access information regarding broadband availability

---

[16] *Inquiry Concerning Deployment of Advanced Telecommunications Capability to All Americans in a Reasonable and Timely Fashion,* 2019 Broadband Deployment Report, FCC 19-44 (rel. May 29, 2019) at p. 2.

at a specific location. Ideally, this lookup tool would reveal multiple data points at an address, including, at a minimum, availability, crowdsourced usage/speed tests, and subscription data. By enriching the lookup tool with these data points, the public and other stakeholders will be able to identify discrepancies in the data being collected. In our experience, data that is inspected and used is of higher quality.

In sum, Microsoft fully supports establishing a broadband fabric as well as the addition of a lookup tool. By taking the actions discussed in this section, the Commission will improve data consistency and the public will be able to use commercially available products, including satellite imagery, to validate the submitted data.

V. Microsoft Supports Several Improvements to the DODC Process.

The Commission seeks comment on a variety of issues relating to how broadband data is submitted into the DODC.[17]

To enable more granularity than the current census block level reporting, submitted polygons should be no smaller than the building level, and in the aggregate, every building in the United States should be represented. Moreover, all polygons must be anchored onto the same fabric and use the same geocoding reference points so that a single standard base map is used to count serviceable locations. Finally, all maps and data sets must contain uniform characteristics. These elements are essential to enabling every person analyzing the data to be on the same page at all times.

---

[17] *Second Notice*, at ¶¶ 78-84.

VI. The Commission Should Ensure Accountability for Filers of Fixed Broadband Data.

The *Second Notice* seeks comment on "the best method (or mix of methods) to ensure the submission of accurate fixed broadband deployment data, including the plans that USAC must develop for corroborating and spot-checking data submitted by fixed providers."[18]

At the outset, adopting data collection best practices recommended in Section II above will promote accountability. By using machine learning tools that incorporate a variety of data sets and making data public for review and improvement, the Commission will increase the quality of its broadband availability data and curb the submission of incorrect data by providers in the first place.

Microsoft is sensitive to burdens placed on providers, especially small providers, in connection with reporting broadband availability data to the Commission. Indeed, Microsoft has advocated for improved FCC Form 477 instructions as a means of simplifying the task for filers and ultimately improving the quality of submitted data.[19] Accordingly, Microsoft does not support penalties for filers that in good faith submit data that proves to be inaccurate.

That said, Microsoft does not believe that reporting fixed broadband data is a complex task for providers that have been complying with FCC Form 477 for many years. Moreover, the *Second Notice* suggests that the Commission will conduct educational outreach to assist filers,

---

[18] *Id.* at ¶ 83.

[19] *See, e.g., Petition for Reconsideration* filed by Microsoft Corporation in WC Docket 19-195 and WC Docket No. 11-10 (Sept 23, 2019) at https://ecfsapi.fcc.gov/file/109241260214870/2019%200923%20Microsoft%20Petition%20for%20Reconsideration%20-%20FINAL%20As%20Filed.pdf.

especially small businesses, with the new DODC process. Once the DODC process is in place, there must be some level of accountability, and the possibility of penalties for providers recklessly or intentionally submitting inaccurate fixed broadband mapping data.

As recognized by legislation introduced this year, the willful or intentional act of significantly overstating broadband coverage frustrates the Commission's goal of accurately targeting resources where they are needed, and may in some cases be anticompetitive.[20] Accordingly, the prospect of a monetary forfeiture can act as a significant deterrent to intentional or reckless actions, will increase accuracy at the outset, and should limit the need for enforcement actions later.

VII. Conclusion.

Adopting the DODC sets the Commission on a course to more accurately measure broadband throughout the nation. Machine learning tools, incorporating rich data sources such as subscription data, usage data, availability data, and demographic data, will enable the Commission to expose potential quality issues in broadband availability data in new and more efficient ways. By implementing these improvements, along with others suggested above, the Commission will be able to make data-driven policy choices that can help to deliver telecommunications and information services to rural consumers that are reasonably

[20] *See, e.g.,* The Broadband Deployment Accuracy and Technological Availability Act, S.1822 (116th Cong.) at Sec. 4, at https://www.congress.gov/bill/116th-congress/senate-bill/1822/text.

comparable in quality and in price to those available in urban areas, the goal set forth by Congress in the 1996 Act.[21]

Respectfully submitted,

MICROSOFT CORPORATION

Paula Boyd, Senior Director
U.S. Government and Regulatory Affairs
901 K Street, N.W., 11th Floor
Washington, DC 20001
202-263-5946

David A. LaFuria
Lukas, LaFuria, Gutierrez & Sachs, LLP
8300 Greensboro Drive, Suite 1200
Tysons, VA 22102
703-584-8678

October 7, 2019

[21] *See* 47 U.S.C. § 254(b)(3).

Exhibit A

Microsoft Notice of Ex Parte Presentation

September 20, 2019

Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5900
Fax 202-783-0583
http://www.microsoft.com/

Microsoft

September 20, 2019

VIA ELECTRONIC FILING

Marlene H. Dortch, Secretary
Federal Communications Commission
445 Twelfth Street, S.W.
Washington, DC 20554

Re: WC Docket No. 19-195
WC Docket No. 11-10

Madam Secretary:

In accordance with Section 1.1206(b) of the Commission's rules,[1] this letter provides notice of oral ex parte presentations to the Commission in the above-captioned dockets. On September 18, 2019, undersigned counsel, along with Paula Boyd, Paul Garnett, John Kahan, and Allen Kim of Microsoft Corporation, met with Allison Baker, Kirk Burgee, Joseph Calascione, Travis Litman, Steven Rosenberg, Arielle Roth, and Alisa Valentin.

In its presentations, Microsoft discussed several aspects of the Report and Order and Second Further Notice of Proposed Rulemaking adopted by the Commission in the above-captioned dockets,[2] as well other related issues, including Microsoft's development of a Machine Learning Model ("ML Model") that it believes can assist the Commission as well as industry and other stakeholders in detecting potential outliers that diverge from reported broadband[3] availability information.

Microsoft commended the Commission for its adoption of the *Report and Order* and for the issues the Commission has raised in the *Further Notice*, stating that it is encouraged by the steps the Commission has taken to improve the accuracy and granularity of broadband

[1] 47 C.F.R. § 1.1206(b).

[2] *Establishing the Digital Opportunity Data Collection*, *et al*., WC Docket No. 19-195, *et al*., Report and Order and Second Further Notice of Proposed Rulemaking, FCC 19-79 (rel. Aug. 6, 2019) ("*Report and Order*" or "*Further Notice*").

[3] Unless otherwise stated herein, "broadband" is intended to refer to broadband service providing 25 Mbps download and 3 Mbps upload speeds.

mapping data, and by the further actions it plans to pursue in its efforts to track the availability and usage of broadband services. Several topics addressed during Microsoft's meetings with Commission staff are discussed in the following paragraphs.

1. FCC Form 477 and Digital Opportunity Data Collection Reporting.

Microsoft praised the decision in the *Report and Order* that, for purposes of the Digital Opportunity Data Collection ("DODC"), broadband service will be treated as actually available "if the reporting fixed service provider has a current broadband connection or it could provide such a connection *within 10 business days* of a customer request, without an extraordinary commitment of resources, and *without construction charges or fees exceeding an ordinary service activation fee*."[4] Microsoft observed that the 10-day criterion, and the restriction placed on construction charges and fees, are significant improvements, that will enable greater precision in determining whether broadband service is "available," and that will help to reduce overstated coverage in broadband availability reports.

Microsoft expressed concern that the Commission opted in the *Report and Order* not to make any revisions to the FCC Form 477 ("Form 477") instructions concerning reporting the availability of fixed broadband connections in a census block. As a result, the Form 477 and the reporting criteria for the DODC are now misaligned, and without any improvements to Form 477, difficulties remain in accurately determining broadband availability in the near term. Microsoft advocated that the Commission should act expeditiously to improve the accuracy of Form 477 by correcting this misalignment.

Microsoft suggested that the best way to improve the accuracy of Form 477 would be to revise Section 5.3 of the Form 477 Instructions. The Instructions currently provide that a fixed broadband connection is considered "available" if the service provider "could, within a service interval that is typical for that type of connection—that is, without an extraordinary commitment of resources—provision" broadband service[5] Microsoft observed that this loose formulation has the disadvantage of leaving considerable discretion to the reporting service provider, and consequently has contributed to the overstatement of broadband availability. Revising the Section 5.3 formulation of availability—to mirror the new DODC criteria—would help to reduce the risk of overstated broadband coverage being reported by Form 477 filers and cure the misalignment between the Form 477 and DODC definitions.

Microsoft also advocated that the Commission should act quickly to align the Form 477 and DODC criteria for determining broadband availability, as a first step in the transition from the use of Form 477 data to the use of DODC data.

---

[4] *Report and Order* at para. 13 (emphasis added) (footnote omitted).

[5] FCC Form 477, Local Telephone Competition and Broadband Reporting, Instructions, at § 5.3.

2. Using a Machine Learning Model to Identify Potential Outlier ZIP Codes.

A principal focus of Microsoft's presentations was a discussion of its development of an ML Model to predict broadband service availability, and a review of data it has generated through the use of this model. Microsoft has used its model to identify potential outlier ZIP Codes, based on comparisons with Form 477 broadband coverage data submitted to the Commission.[6]

In developing the ML Model, Microsoft's objective has been to design a means by which machine learning capabilities can be utilized to identify potential inaccuracies in reported broadband availability data, and thus warrant further examination by the Commission, or by industry or other stakeholders. The model is highly accurate at the national level, predicting broadband availability with 90 percent accuracy. The model's accuracy diminishes somewhat at the more granular ZIP Code level, demonstrating a capability to predict 78 percent of ZIP Codes within 10 percentage points of reported broadband availability, and a capability to predict 63 percent of ZIP Codes within 5 percentage points of reported broadband availability. These levels of accuracy at the ZIP Code level are sufficient, however, to make the ML Model a useful tool in identifying ZIP Codes that may have inaccuracies in reported broadband availability.

The ML Model is built to use usage data and demographic data to enable predictions of broadband availability, thus serving as a useful auditing and crowdsourcing tool in identifying where there may be problems with availability and broadband service quality. Microsoft utilized an extensive array of device-level data accumulated from more than 200 Microsoft services[7] as a basis for estimating broadband download speeds and broadband service coverage. Microsoft indicated that its usage data is extremely precise, showing the "ground truth" concerning progress being made in the adoption of broadband services across the nation.

---

[6] Microsoft used availability data reported to the Commission at the census tract level and "converted" this data to illustrate availability on a ZIP Code basis, using Department of Housing and Urban Development ("HUD") USPS Zip Code Crosswalk Files. The HUD files were used by Microsoft to determine the percentage of each census tract that is associated with a particular ZIP Code. Microsoft chose to analyze its ML Model data on a ZIP Code basis because that is the lowest level of granularity for which Microsoft has confidence in the data generated by the model, and because consumers are more familiar with ZIP Codes than with census tract numbers.

[7] Microsoft did not utilize any personally identifiable information ("PII") in connection with its design and use of the ML Model.

Combined with various other demographic indicators (*e.g.*, land devoted to agriculture, education levels, median household income, home ownership),[8] this usage data works as a powerful predictor of broadband availability.

Microsoft noted that, currently, there are discrepancies concerning nationwide data relating to broadband availability and broadband usage. Data reported to the Commission by service providers indicates that 93.5 percent of all areas in the nation have access to fixed broadband service, while Microsoft estimates, on the basis of its usage data, that approximately 49 percent of all Americans are accessing the internet at broadband speeds. Microsoft observed that, although availability can be expected to be somewhat "north" of usage, such a substantial gap suggests issues with the availability data.

The Commission's subscriber data and other customer survey data correlate relatively closely with Microsoft's usage data, thus confirming Microsoft's view that the discrepancies between broadband availability and usage data are not caused by inaccurate usage data. Microsoft indicated in its presentations that the Commission's data shows a nationwide broadband subscribership level of 54 percent, that a Pew Research survey shows that 65 percent of respondents indicate that they are home broadband users, and that consumer responses gathered by American FactFinder (Census Bureau) show that 67 percent of consumers have broadband subscriptions.

Microsoft explained to Commission staff that it has tested the utility of its ML Model by using it to identify potential outlier ZIP Codes. Microsoft presented the top 20 ZIP Codes that its ML Model indicate are likely the most challenged with respect to broadband availability, and noted that the 20 ZIP Codes also show substantial gaps between estimated broadband availability (based on data reported to the Commission) and the estimated percentage of consumer broadband usage (based on Microsoft's data). Microsoft also presented 12-month speed test charts (from BroadbandNow.com) for each of the 20 outlier ZIP Codes, which show that, for almost all of the months for all of the 20 ZIP Codes, available broadband speeds were below 25/3 Mbps, calling into question the extent to which broadband actually is available in any of these ZIP Codes and confirming the results of the ML Model.

Microsoft stated that the results produced by its model simply validate what we already know: Broadband availability data reported to the Commission tends to overstate availability. Microsoft expressed the view that the model, especially when combined with subscription data (see below), may serve the Commission, as well as industry and other stakeholders, as a useful auditing tool to identify where the broadband availability problems are.

[8] Among the indicators used, usage, not surprisingly, is the most important variable in predicting availability, by a considerable margin.

Microsoft indicated in its presentations that it plans to make both its ML Model, and data the model has produced, available to the public. It plans to make the model open-sourced and publicly available on GitHub, enabling all stakeholders to utilize the model, as well as giving them an opportunity to strengthen the model's performance by adding their own datasets to the model. As Microsoft continues to use the model to identify outlier ZIP Codes, it also plans to publish additional ZIP Code results after confirming those results.

3. Pairing Broadband Coverage Polygons/Shapefiles with Datasets.

The *Report and Order* requires all fixed broadband providers to submit broadband coverage polygons depicting the areas where the providers actually have broadband-capable networks and make fixed broadband service available to end-user customers.[9] Microsoft expressed its support for this decision to require the submission of polygons, in part because the use of broadband coverage polygons will cure one of the problems, the lack of granularity, associated with current broadband availability reporting criteria.

Specifically, service providers currently report lists of census blocks in which they provide broadband service, but they are not required to report the *extent of their coverage* within a census block. Thus, if they serve only one location in a census block, the Commission treats the entire census block as being served. Now, broadband providers will be required to certify that broadband is available at every location within their reported polygons.

Microsoft also addressed the issue of what datasets should be employed in conjunction with the polygon maps, advocating that individual buildings are likely best as the locations that must have access to broadband. Microsoft acknowledged that the use of a buildings/structures dataset will require the resolution of various issues (*e.g.*, whether and how to distinguish between residential and commercial buildings).

4. Crowdsourcing.

Microsoft expressed its view that crowdsourced data serves as another effective tool for verifying broadband providers' service availability reports and expressed its support for the Commission's requiring the use of crowdsourcing. Microsoft suggested that the Commission should collect crowdsourced data that relates to both broadband availability and broadband usage. Microsoft noted that, based on its work with its ML Model, usage data is the most critical input into identifying where the Commission may have potential outlier areas with respect to its broadband availability dataset. Without usage data, it could be very difficult to audit the Commission's availability data effectively.

[9] *Report and Order* at para. 12.

5. The Use of Subscription Data.

Microsoft expressed the view that the granular 25/3 Mbps broadband subscription data collected by the Commission should be made public, to the extent practicable. In general, subscription data can serve as an effective means of validating the accuracy of reported broadband availability. More specifically, the Commission's subscription data, if made publicly available, could be used as an additional dataset for Microsoft's ML Model, thus making the model an even more accurate predictor of availability.

* * * * * * *

Copies of Microsoft's presentation slides are enclosed for the record. The "FCC1" slide deck was used in meetings with Travis Litman and Arielle Roth, the "FCC2" slide deck was used in a meeting with Allison Baker, Kirk Burgee, and Steven Rosenberg, and the "FCC3" slide deck was used in meetings with Joseph Calascione and Alisa Valentin.

Should you have any questions, please contact the undersigned directly.

Sincerely,

MICROSOFT CORPORATION

Paula Boyd, Senior Director
U.S. Government and Regulatory Affairs
901 K Street, N.W., 11th Floor
Washington, DC 20001
202-263-5946

John Cimko
Lukas, LaFuria, Gutierrez & Sachs, LLP
8300 Greensboro Drive, Suite 1200
Tysons, VA 22102
703-584-8686

Counsel to Microsoft Corporation

Enclosures

cc: Allison Baker
Kirk Burgee
Joseph Calascione
Travis Litman

Steven Rosenberg
Arielle Roth
Alisa Valentin
Paul Garnett
John Kahan
Allen Kim